ABSTRACT

The present invention relates to a method for determining relationships among a protein's biochemical properties, biophysical properties and amino acid sequence is provided. The invention further provides a database of protein sequence information and experimentally determined protein properties. This database is analyzed using data-mining techniques to find correlations among protein sequence information, biochemical properties and biophysical properties. Using the empirical correlations obtained from the data-mining techniques, the properties of new proteins are determined given their amino acid sequence information alone or using a combination of the sequence information and one or more properties.